Individual Privacy Accounting via a Rényi Filter

Vitaly Feldman
Apple
vitaly.edu@gmail.com

Tijana Zrnic*
University of California, Berkeley
tijana.zrnic@berkeley.edu

Abstract

We consider a sequential setting in which a single dataset of individuals is used to perform adaptively-chosen analyses, while ensuring that the differential privacy loss of each participant does not exceed a pre-specified privacy budget. The standard approach to this problem relies on bounding a worst-case estimate of the privacy loss over all individuals and all possible values of their data, for every single analysis. Yet, in many scenarios this approach is overly conservative, especially for "typical" data points which incur little privacy loss by participation in most of the analyses. In this work, we give a method for tighter privacy loss accounting based on the value of a personalized privacy loss estimate for each individual in each analysis. The accounting method relies on a new composition theorem for Rényi differential privacy, which allows adaptively-chosen privacy parameters. We apply our results to the analysis of noisy gradient descent and show how existing algorithms can be generalized to incorporate individual privacy accounting and thus achieve a better privacy-utility tradeoff.

1 Introduction

Understanding how privacy of an individual degrades as the number of analyses using their data grows is of paramount importance in privacy-preserving data analysis. On one hand, this allows individuals to participate in multiple statistical analyses, all the while knowing that their privacy cannot be compromised by aggregating the resulting reports. On the other hand, this feature is crucial for private algorithm design — instead of having to reason about the privacy properties of a complex algorithm, it allows reasoning about the privacy of the subroutines that make up the final algorithm.

For differential privacy [9], this accounting of privacy losses is typically done using composition theorems. Importantly, given that statistical analyses often rely on the outputs of previous analyses, and that algorithmic subroutines feed into one another, the composition theorems need to be *adaptive*, namely, allow the choice of which algorithm to run next to depend on the outputs of all previous computations. For example, in gradient descent, the computation of the gradient depends on the value of the current iterate, which itself is the output of the previous steps of the algorithm.

Given the central role that adaptive composition theorems play for differentially private data analysis, they have been investigated in numerous works (e.g. [10, 16, 8, 20, 19, 3, 22, 5, 23]). While they differ in some aspects, they also share one limitation. Namely, all of these theorems reason about the worst-case privacy loss for each constituent algorithm in the composition. Here, "worst-case" refers to the worst choice of individual in the dataset and worst choice of value for their data. This pessimistic accounting implies that every algorithm is summarized via a single privacy parameter, shared among all participants in the analysis.

In most scenarios, however, different individuals have different effects on each of the algorithms, as measured by differential privacy. More precisely, the output of an analysis may have little to no

*Work done while at Apple.
Full paper available at: https://arxiv.org/pdf/2008.11193.pdf

34th Conference on Neural Information Processing Systems (NeurIPS 2020), Vancouver, Canada.

dependence on the presence of some individuals. For example, if we wish to report the average income in a neighborhood, removing an individual whose income is close to the average has virtually no impact on the final report after noise addition. Similarly, when training a machine learning model via gradient descent, the norm of the gradient defined by a data point is often much smaller than the maximum norm (typically determined by a clipping operation). As a result, in many cases no single individual is likely to have the worst-case effect on all the steps of the analysis. This means that accounting based on existing composition theorems may be unnecessarily conservative.

We present a tighter analysis of privacy loss composition by computing the associated divergences at an individual level. In particular, to achieve a pre-specified privacy budget, we keep track of a personalized estimate of the divergence for each individual in the analyzed dataset and ensure that the respective estimate for all individuals is maintained under the budget throughout the composition.

1.1 Preliminaries

We let $S = (X_1, \ldots, X_n)$ denote the analyzed dataset, and $S^{-i} \stackrel{\text{def}}{=} (X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n)$ the analyzed dataset after removing point X_i. If the algorithm requires an input of fixed size, one can obtain the same results for algorithms that replace X_i with an arbitrary fixed element $X^\star$ (e.g. 0).

Our analysis makes use of Rényi differential privacy (RDP), a relaxation of differential privacy (DP) based on Rényi divergences which often leads to tighter privacy bounds than analyzing DP directly. For two random variables X, Y, we denote by $D_\alpha(X\|Y)$ the Rényi divergence of order α between the distribution of X and the distribution of Y.

Definition 1.1 ([19])**.** *A randomized algorithm $\mathcal{A}$ is (α, ρ)-Rényi differentially private (RDP) if for all datasets $S = (X_1, \ldots, X_n)$ and $i \in [n]$, it holds that* $\max\{D_\alpha\left(\mathcal{A}(S)\|\mathcal{A}(S^{-i})\right), D_\alpha\left(\mathcal{A}(S^{-i})\|\mathcal{A}(S)\right)\} \leq \rho$.

Although our guarantees will be stated in terms of RDP, RDP can easily be converted to DP [19].

Our main object of study is *adaptive composition*. Here, for a given input dataset S and a sequence of algorithms $(\mathcal{A}_t)_{t=1}^k$, one sequentially computes reports $a_t = \mathcal{A}_t(a_1, \ldots, a_{t-1}, S)$ as a function of previous reports and the input dataset. We denote by $\mathcal{A}^{(k)}(S) \stackrel{\text{def}}{=} (a_1, \ldots, a_k)$ the output of k steps of adaptive composition. If $\mathcal{A}_t(a_1, \ldots, a_{t-1}, \cdot)$ is (α, ρ_t)-RDP for *all* values of $a_1, \ldots, a_{t-1}$, then the standard adaptive composition for RDP says that $\mathcal{A}^{(k)}$ is $(\alpha, \sum_{t=1}^k \rho_t)$-RDP [19]. Note that the parameters $\rho_1, \ldots, \rho_k$ are independent of the specific reports $a_1, \ldots, a_k$ obtained in the composition.

Our individual accounting relies on an RDP version of personalized differential privacy [13].

Definition 1.2 (Individual RDP)**.** *Fix $n \in \mathbb{N}$ and a data point X. We say that a randomized algorithm $\mathcal{A}$ satisfies* (α, ρ)-individual Rényi differential privacy for X *if for all datasets $S = (X_1, \ldots, X_m)$ such that $m \leq n$ and $X_i = X$ for some i, it holds that* $\max\{D_\alpha\left(\mathcal{A}(S)\|\mathcal{A}(S^{-i})\right), D_\alpha\left(\mathcal{A}(S^{-i})\|\mathcal{A}(S)\right)\} \leq \rho$.

2 Individual privacy filtering

It is straightforward to measure the worst-case effect of a specific data point on a given analysis: one can simply replace the supremum over all datasets in the standard definition of (removal) differential privacy with the supremum over datasets that include that specific data point (as in Definition 1.2). Indeed, such a definition was given in the work of Ebadi et al. [13] and a related definition is given by Wang [24]. However, a meaningful application of adaptive composition with such a definition immediately runs into the following technical challenge. Standard adaptive composition theorems require that the privacy parameter of each step be fixed in advance. For individual privacy parameters, this approach requires using the worst-case value of the individual privacy loss over all the possible analyses at a given step. Individual privacy parameters tend to be much more sensitive to the analysis being performed than worst-case privacy losses, and thus using the worst-case value over all analyses is likely to negate the benefits of using individual privacy losses in the first place.

Thus the main technical challenge in analyzing composition of individual privacy losses is that they are themselves random variables that depend on the outputs of all the previous computations. More specifically, if we denote by $a_1, \ldots, a_{t-1}$ the output of the first $t-1$ adaptively composed algorithms $\mathcal{A}_1, \ldots, \mathcal{A}_{t-1}$, then the individual privacy loss of any point incurred by applying algorithm $\mathcal{A}_t$ is a

function of $a_1, \ldots, a_{t-1}$. Therefore, to tackle the problem of composing individual privacy losses we need to understand composition with *adaptively-chosen* privacy parameters in general.

This setting is rather subtle and even defining privacy in terms of adaptively-chosen privacy parameters requires some care. This setting was first studied by Rogers et al. [22], who introduced the notion of a *privacy filter*. Informally, a privacy filter is a stopping time rule that halts a computation based on the adaptive sequence of privacy parameters and ensures that a pre-specified privacy budget is not exceeded. Rogers et al. define a filter for approximate differential privacy that asymptotically behaves like the advanced composition theorem [10], but is substantially more involved and loses a constant factor. Moreover, several of the tighter analyses of Gaussian noise addition require composition to be done in Rényi differential privacy [1, 19].

Our main result can be seen as a privacy filter for Rényi differential privacy (RDP) whose stopping rule exactly matches the rate of standard RDP composition [19].

Theorem 2.1. *Fix any $B \geq 0, \alpha \geq 1$. Suppose that $\mathcal{A}_t$ is (α, ρ_t)-Rényi differentially private, where ρ_t is an arbitrary function of $a_1, \ldots, a_{t-1}$. If $\sum_{t=1}^{k} \rho_t \leq B$ holds almost surely, then the adaptive composition of $\mathcal{A}_1, \ldots, \mathcal{A}_k$ is (α, B)-Rényi differentially private.*

This theorem implies that stopping the analysis based on the sum of privacy parameters so far is valid even with adaptive privacy parameters. Our RDP filter immediately implies a filter for approximate differential privacy that is as tight as any version of the advanced composition theorem obtained via concentrated differential privacy [3]. These Rényi-divergence-based composition analyses are known to improve upon the rate of Dwork et al. [10] and, in particular, improve on the results in [22].

Theorem 2.1 allows us to define an *individual privacy filter*, which adaptively drops points from the analysis once their *personalized* privacy loss estimate exceeds the privacy budget. At every step t, the filter determines an active set of points $S_t \subseteq S$ based on cumulative estimated losses, and applies $\mathcal{A}_t$ only to S_t. We let $\rho_t^{(i)}$ denote the individual RDP parameter of order α of $\mathcal{A}_t$ for point X_i.

Algorithm 1 Adaptive composition with individual privacy filtering

input : dataset $S \in \mathcal{X}^n$, sequence of algorithms $\mathcal{A}_t$, $t = 1, 2, \ldots, k$
for $t = 1, \ldots, k$ **do**
  For all $i \in [n]$, compute $\rho_t^{(i)}$
  Determine active set $S_t = (X_i \ : \ \sum_{j=1}^{t} \rho_j^{(i)} \leq B)$
  For all $i \in [n]$, set $\rho_t^{(i)} \leftarrow \rho_t^{(i)} \mathbf{1}\{X_i \in S_t\}$
  Compute $a_t = \mathcal{A}_t(a_1, \ldots, a_{t-1}, S_t)$
end
Return $(a_1, \ldots, a_k)$

Theorem 2.2. *Algorithm 1 satisfies (α, B)-Rényi differential privacy.*

Application to optimization. A popular approach to private model training via gradient descent (GD) is to clip the norm of individual gradients at every time step and add Gaussian noise to the clipped gradients [1]. Existing privacy analyses compute the overall privacy spent up to a given round by using a uniform upper bound on the gradient norms, determined by the clipping value. In contrast, the individual privacy filter allows utilization of data points as long as their *realized* gradients have low norm. We state the generalization of private gradient descent with individual filtering in the full paper version. We compare the accuracies achieved by the two algorithms on the MNIST [17] and Adult [6] datasets, for a fixed (ε, δ)-DP budget. We fix $\delta = 10^{-5}$. In high privacy regimes, namely when ε is small, we observe that individual filtering achieves noticeably higher accuracies, as the standard private GD algorithm misclassifies points which do not use up their gradient budget.

	MNIST accuracies		Adult accuracies	
ε	private GD	private GD w/ filtering	private GD	private GD w/ filtering
0.3	$(92.80 \pm 0.52)\%$	$(93.18 \pm 0.32)\%$	$(83.80 \pm 0.22)\%$	$(83.91 \pm 0.12)\%$
0.5	$(94.62 \pm 0.43)\%$	$(94.90 \pm 0.26)\%$	$(84.11 \pm 0.15)\%$	$(84.18 \pm 0.10)\%$
1.2	$(96.56 \pm 0.15)\%$	$(96.56 \pm 0.15)\%$	$(84.45 \pm 0.13)\%$	$(84.48 \pm 0.15)\%$

Application to privacy odometers. Rogers et al. [22] define a *privacy odometer*, which provides an upper bound on the running privacy loss, and does not require a pre-specified budget. By applying a discretization argument, we make an observation that our Rényi privacy filter can be utilized to design an approximate *personalized* odometer of an individual's own privacy loss.

Algorithm 2 Individual privacy loss tracking via Rényi privacy odometer

input : dataset $S \in \mathcal{X}^n$, discretization error $\Delta > 0$, sequence of algorithms $(\mathcal{A}_t)_{t=1}^k$, index $i \in [n]$

Initialize odometer $O_1^{(i)} = \Delta$ and set $k = 1, T_0^{(i)} = 1$

for $t = 1, 2, \ldots$ **do**

  Compute $a_t = \mathcal{A}_t(a_1, \ldots, a_{t-1}, S)$, set $O_t^{(i)} \leftarrow O_{t-1}^{(i)}$

  if $\sum_{j=T_{k-1}^{(i)}}^{t} \rho_j^{(i)} > \Delta$ **then**

    Augment odometer $O_t^{(i)} \leftarrow O_t^{(i)} + \Delta$, set $T_k^{(i)} \leftarrow t, k \leftarrow k+1$

  end

end

Corollary 2.3. *Fix $i \in [n]$, and suppose that $\rho_j^{(i)} \leq \Delta$ almost surely, for all $j \in \mathbb{N}$. Let t be any time such that $T_{k-1}^{(i)} \leq t < T_k^{(i)}$. Then, $O_t^{(i)}$ upper bounds the individual privacy loss of point X_i at time t:* $\max\{D_\alpha\left(\mathcal{A}^{(t)}(S) \| \mathcal{A}^{(t)}(S^{-i})\right), D_\alpha\left(\mathcal{A}^{(t)}(S^{-i}) \| \mathcal{A}^{(t)}(S)\right)\} \leq k\Delta = O_t^{(i)}$.

3 Related work

The main motivation behind our work is obtaining tighter privacy accounting methods through, broadly speaking, "personalized" accounting of privacy losses. Existing literature in differential privacy discusses several related notions [15, 13, 24, 4], although typically with an incomparable objective. Ghosh and Roth [15] discuss individual privacy in the context of selling privacy at auction and their definition does not depend on the value of the data point but only on its index in the dataset. Cummings and Durfee [4] rely on a similar privacy definition, investigate an associated definition of individual sensitivity, and demonstrate a general way to preprocess an arbitrary function of a dataset into a function that has the desired bounds on individual sensitivities.

Ebadi et al. [13] introduce personalized differential privacy in the context of private database queries and describe a system which drops points when their personalized privacy loss exceeds a budget. While this type of accounting is similar to ours in spirit, their work only considers basic and non-adaptive composition. The work of Wang [24] considers the privacy loss of a specific data point relative to a fixed dataset. It provides techniques for evaluating this "per-instance" privacy loss for several statistical problems. Wang [24] also briefly discusses adaptive composition of per-instance differential privacy as a straightforward generalization of the usual advanced composition theorem [10], but the per-instance privacy parameters are assumed to be *fixed*. As discussed above, having fixed per-instance privacy parameters, while allowing adaptive composition, is likely to negate some of the benefits of personalized privacy estimates. The work of Ligett et al. [18] tightens individuals' personalized privacy loss by taking into account subsets of analyses in which an individual does not participate. Within the studies in which an individual participates, however, they consider the usual worst-case privacy loss, rather than an individual one. In addition, the analyses in which the user participates are determined in a data-independent way.

Our work can be seen as related to data-dependent approaches to analyses of DP algorithms such as smooth sensitivity [21], the propose-test-release framework [7], and *ex-post* privacy guarantees [25]. The focus of our work is complementary in that we aim to capture the dependence of the output on the value of each individual's data point as opposed to the "easiness" of the entire dataset. Our approach also requires composition to exploit the gains from individual privacy loss accounting.

Finally, adaptive composition of differentially private algorithms is one of the key tools for establishing statistical validity of an adaptively-chosen sequence of statistical analyses [12, 11, 2]. In this context, Feldman and Steinke [14] show that the individual KL-divergence losses (or RDP losses for $\alpha = 1$) compose adaptively and can be exploited for deriving tighter generalization results. However, their results still require that the average of individual KL-divergences be upper bounded by a fixed worst-case value and the analysis appears to be limited to the $\alpha = 1$ case.

References

[1] Martin Abadi, Andy Chu, Ian Goodfellow, H Brendan McMahan, Ilya Mironov, Kunal Talwar, and Li Zhang. Deep learning with differential privacy. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, pages 308–318, 2016.

[2] Raef Bassily, Kobbi Nissim, Adam Smith, Thomas Steinke, Uri Stemmer, and Jonathan Ullman. Algorithmic stability for adaptive data analysis. In *Proceedings of the forty-eighth annual ACM symposium on Theory of Computing*, pages 1046–1059, 2016.

[3] Mark Bun and Thomas Steinke. Concentrated differential privacy: Simplifications, extensions, and lower bounds. In *Theory of Cryptography Conference*, pages 635–658. Springer, 2016.

[4] Rachel Cummings and David Durfee. Individual sensitivity preprocessing for data privacy. In *Proceedings of the Fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 528–547. SIAM, 2020.

[5] Jinshuo Dong, Aaron Roth, and Weijie J Su. Gaussian differential privacy. *arXiv preprint arXiv:1905.02383*, 2019.

[6] Dheeru Dua and Casey Graff. UCI machine learning repository, 2017. URL `http://archive.ics.uci.edu/ml`.

[7] Cynthia Dwork and Jing Lei. Differential privacy and robust statistics. In *Proceedings of the forty-first annual ACM symposium on Theory of computing*, pages 371–380, 2009.

[8] Cynthia Dwork and Guy N Rothblum. Concentrated differential privacy. *arXiv preprint arXiv:1603.01887*, 2016.

[9] Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. Calibrating noise to sensitivity in private data analysis. In *Proc. of the Third Conf. on Theory of Cryptography (TCC)*, pages 265–284, 2006. URL `http://dx.doi.org/10.1007/11681878_14`.

[10] Cynthia Dwork, Guy N Rothblum, and Salil Vadhan. Boosting and differential privacy. In *2010 IEEE 51st Annual Symposium on Foundations of Computer Science*, pages 51–60. IEEE, 2010.

[11] Cynthia Dwork, Vitaly Feldman, Moritz Hardt, Toni Pitassi, Omer Reingold, and Aaron Roth. Generalization in adaptive data analysis and holdout reuse. In *Advances in Neural Information Processing Systems*, pages 2350–2358, 2015.

[12] Cynthia Dwork, Vitaly Feldman, Moritz Hardt, Toniann Pitassi, Omer Reingold, and Aaron Leon Roth. Preserving statistical validity in adaptive data analysis. In *Proceedings of the forty-seventh annual ACM symposium on Theory of computing*, pages 117–126, 2015.

[13] Hamid Ebadi, David Sands, and Gerardo Schneider. Differential privacy: Now it's getting personal. *Acm Sigplan Notices*, 50(1):69–81, 2015.

[14] Vitaly Feldman and Thomas Steinke. Calibrating noise to variance in adaptive data analysis. In *Conference On Learning Theory*, pages 535–544, 2018.

[15] Arpita Ghosh and Aaron Roth. Selling privacy at auction. In *Proceedings of the 12th ACM conference on Electronic commerce*, pages 199–208, 2011.

[16] Peter Kairouz, Sewoong Oh, and Pramod Viswanath. The composition theorem for differential privacy. *IEEE Transactions on Information Theory*, 63(6):4037–4049, 2017.

[17] Yann LeCun, Corinna Cortes, and CJ Burges. Mnist handwritten digit database. *ATT Labs [Online]. Available: http://yann.lecun.com/exdb/mnist*, 2, 2010.

[18] Katrina Ligett, Charlotte Peale, and Omer Reingold. Bounded-leakage differential privacy. In *1st Symposium on Foundations of Responsible Computing (FORC 2020)*. Schloss Dagstuhl-Leibniz-Zentrum für Informatik, 2020.

[19] Ilya Mironov. Rényi differential privacy. In *2017 IEEE 30th Computer Security Foundations Symposium (CSF)*, pages 263–275. IEEE, 2017.

[20] Jack Murtagh and Salil Vadhan. The complexity of computing the optimal composition of differential privacy. In *Theory of Cryptography Conference*, pages 157–175. Springer, 2016.

[21] Kobbi Nissim, Sofya Raskhodnikova, and Adam Smith. Smooth sensitivity and sampling in private data analysis. In *Proceedings of the thirty-ninth annual ACM symposium on Theory of computing*, pages 75–84, 2007.

[22] Ryan M Rogers, Aaron Roth, Jonathan Ullman, and Salil Vadhan. Privacy odometers and filters: Pay-as-you-go composition. In *Advances in Neural Information Processing Systems*, pages 1921–1929, 2016.

[23] David M Sommer, Sebastian Meiser, and Esfandiar Mohammadi. Privacy loss classes: The central limit theorem in differential privacy. *Proceedings on privacy enhancing technologies*, 2019(2):245–269, 2019.

[24] Yu-Xiang Wang. Per-instance differential privacy. *Journal of Privacy and Confidentiality*, 9(1), 2019.

[25] Steven Wu, Aaron Roth, Katrina Ligett, Bo Waggoner, and Seth Neel. Accuracy first: Selecting a differential privacy level for accuracy-constrained ERM. *Journal of Privacy and Confidentiality*, 9(2), 2019.